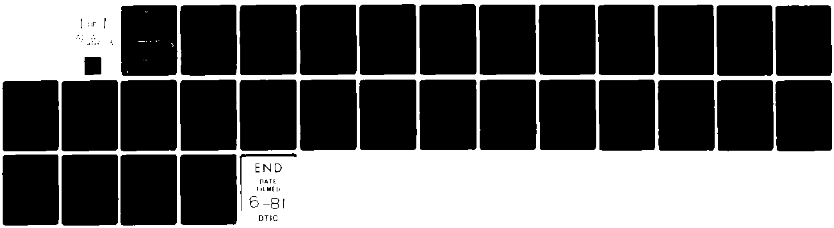
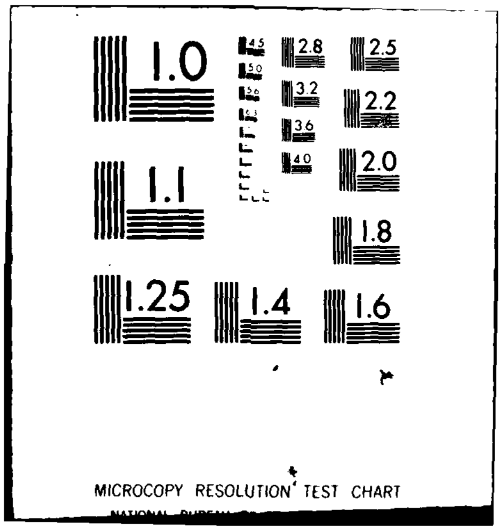
AD-A099 613 PURDUE UNIV LAFAYETTE IN DEPT OF STATISTICS F/G 12/1
ON THE PROBLEM OF FINDING A BEST POPULATION WITH RESPECT TO A C--ETC(U)
MAR 81 S S GUPTA, K MIESCKE N00014-75-C-0455
UNCLASSIFIED MIMEOGRAPH-SER-81-6 NL




1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
MICROCOPY RESOLUTION TEST CHART

LEVEL II

12

PURDUE UNIVERSITY

DEPARTMENT OF STATISTICS

DIVISION OF MATHEMATICAL SCIENCES

(12)

(6) On the Problem of Finding a Best Population
with Respect to a Control in Two Stages,

by

(10) Shanti S. Gupta [redacted] Klaus-J. Miescke
Purdue University and University of Mainz

(14) MIMEOGRAPH-SER-81-6

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #81-6

(11) Mar[redacted]981 (12) 29

*This research was supported by the Office of Naval Research contract
(15) N00014-75-C-0455 at Purdue University. Reproduction in whole or in
part is permitted for any purpose of the United States Government.

ERRATA SHEET

Purdue University
Department of Statistics
Mimeograph Series #81-6

Page 5 line 18↓: replace $\underline{u}$ by $\underline{U}$.

Page 9 line 4↓: replace $\underline{u}$ by $\underline{U}$ and $\underline{v}$ by $\underline{V}$.

On the Problem of Finding a Best Population with Respect to a Control in Two Stages*

by

Shanti S. Gupta and Klaus-J. Miescke
Purdue University and University of Mainz

SUMMARY

Let $\pi_1,\ldots,\pi_k$ be given populations associated with unknown real parameters $\theta_1,\ldots,\theta_k$. π_i is assumed to be "good" if $\theta_i > \theta_0$, where $\theta_0 \in \mathbb{R}$ is a given control value, $i = 1,\ldots,k$. The goal is to find the "best" population (i.e. that one with the largest parameter), if it is "good", in 2 stages with screening out "bad" populations in the first stage. Consideration is restricted to permutation invariant procedures. It is shown that under MLR and a general invariant loss structure the natural final decisions are optimum. More generally an extension of the "Bahadur-Goodman Theorem" to sequential settings (with and without relation to a control) is derived. If the loss structure consists of the cost for sampling plus the loss for final decision, it is shown that for every symmetric prior there exists a Bayes procedure which selects at the first stage populations according to the largest observations. Natural procedures, which screen out with the UMP test for H: $\theta \le \theta_0$ versus K: $\theta > \theta_0$ at fixed level α, are considered. As an example, all results are studied in the special case of normal populations with unknown means and a common known variance.

AMS 1970 subject classification: Primary 62F07, secondary 62F05, 62F15, 62L99.

Key words and phrases: Multiple comparison with a control, two-stage selection procedures, screening procedures, sequential procedures, Bayesian procedures, optimal selection.

*This research was supported by the Office of Naval Research contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

1. Introduction. Let $\pi_1,\ldots,\pi_k$ be k populations associated with unknown parameters $\theta_1,\ldots,\theta_k \in \Omega \subseteq \mathbb{R}$. Let $\theta_0 \in \Omega$ be a given control value such that every π_i with $\theta_i > \theta_0$ is assumed to be "good", and "bad" otherwise, $i = 1,\ldots,k$. We consider the problem: how to find the best population (i.e. that one associated with the largest parameter) among the good ones (if there is any) in two stages by screening out non-best (or bad) populations in the first stage.

Assume that samples $\{X_{ij}\}_{j=1,\ldots,n}$ and $\{Y_{ij}\}_{j=1,\ldots,m}$ can be drawn from π_i at the first and the second stage, respectively, $i = 1,\ldots,k$, which are mutually independent. Let U_i and V_i be real-valued sufficient statistics for θ_i with respect to these samples which have densities f_{θ_i} and g_{θ_i}, respectively, with respect to the Lebesgue measure on $\mathbb{R}$, $i = 1,\ldots,k$. The families $\{f_\theta\}_{\theta\in\Omega}$ and $\{g_\theta\}_{\theta\in\Omega}$ are assumed to be known and to have monotone non-decreasing likelihood ratios (MLR). Finally, let $W_i = T(U_i,V_i)$ be a real-valued sufficient statistic for θ_i with respect to (U_i,V_i), which has a density h_{θ_i} with respect to the Lebesgue measure on $\mathbb{R}$, where the family $\{h_\theta\}_{\theta\in\Omega}$ also has MLR. For notational convenience, let $\underline{U} = (U_1,\ldots,U_k)$, and let $\underline{V}$, $\underline{W}$ etc. have analogous meaning.

In this paper we will study a certain class of 2-stage procedures (S,d), defined as follows. Let S denote any subset selection procedure based on $\underline{U}$, i.e. $S: \mathbb{R}^k \to \{s | s \subseteq \{1,\ldots,k\}\}$ measurable with respect to Borel sets in $\mathbb{R}^k$, where an empty subset is admitted. S acts as a screening procedure in the first stage. Let $d = \{d_s\}_{s \subseteq \{1,\ldots,k\}}$ with $d_\emptyset = 0$ and $d_{\{i\}} = i$, $i = 1,\ldots,k$. Moreover, for every $s \subseteq \{1,\ldots,k\}$ with size $|s| \geq 2$, let $d_s: \mathbb{R}^k \times \mathbb{R}^k \to s$, where $d_s(\underline{u},\underline{v})$ depends only on variables u_i and v_i with $i \in s$, and where d_s is measurable with respect to the Borel sets in their

joint space $\mathbb{R}^{2|s|}$. d represents the set of final decisions at the first stage and the second stage, respectively. The introduction of the (at the first sight) somewhat complicated looking structure d will prove to be very convenient in the sequel. Now we are ready to define our 2-stage procedures in a concise way.

Definition 1. 2-stage procedure (S,d).

Stage 1: Take observations (i.e. the X-samples) from $\pi_1,\ldots,\pi_k$. Select all populations π_i with $i \in s = S(\underline{U})$. If $s = \emptyset$, stop, and decide $d_\emptyset = 0$ (i.e. "no population is good"). If $s = \{i\}$ for some $i \in \{1,\ldots,k\}$, stop, and decide $d_{\{i\}} = i$ (i.e. "π_i is good and the best one"). If $|s| \geq 2$, proceed to Stage 2.

Stage 2: Take additional observations (i.e. the Y-samples) from all populations π_i with $i \in s$ and make the final decision $d_s(\underline{U},\underline{V})$ (i.e. "π_{i_0} is good and the best one", if $d_s(\underline{U},\underline{V}) = i_0$, say, for some $i_0 \in s$.).

Throughout this paper we will restrict consideration to procedures (S,d) which are completely (i.e. with respect to both, S and d) invariant under permutations of the k populations $\pi_1,\ldots,\pi_k$.

In Section 2 it will be shown that under any reasonable loss structure the optimal final decisions are always the natural ones, i.e. are associated with the largest sufficient statistic among those coming from the populations which still are eligible. This result can be derived from Lehmann's (1966) version of the "Bahadur-Goodman-Theorem". In Section 3 a natural type 2-stage procedure will be studied which screens out in the first stage by means of an UMP-test ("$\theta \leq \theta_0$" versus "$\theta > \theta_0$") at a fixed level, which is applied separately to $U_1,\ldots,U_k$, respectively. Finally, in Section 4 it

will be shown that under a fairly general loss structure (cost for sampling plus loss for final decision) and for every symmetric prior there exists a Bayes 2-stage procedure which is *completely monotone* (i.e. where also the subset selections are made in terms of the largest observations), provided that a certain condition (Assumption (A) or (B)) is satisfied. This result will be derived by a two-fold application of Eaton's (1967) more general version of the "Bahadur-Goodman-Theorem". Throughout the following we shall repeatedly study, as an example, the case of k normal populations with unknown means $\theta_1,\ldots,\theta_k$ and a common known variance $\sigma^2 > 0$.

2. Optimality of the natural final decisions. In this section we assume that a loss structure is given which we will specify only with respect to final decisions, and without reference to the control θ_0. This allows us to state the results in a more general setting including also the non-control ("finding the best population") problems such as those studied by Tamhane and Bechhofer (1979).

Definition 2. Loss structure L.

Let us assume that for every procedure (S,d) subsequent decisions $S = s$ and $d_s = i$, $i \in s$, result in a real-valued loss $L(s,i,\underline{\theta})$ at $\underline{\theta} = (\theta_1,\ldots,\theta_k) \in \Omega^k$, which is integrable and has the following two properties:

(a) L is permutation invariant (i.e. $L(\pi s,\pi i,\pi\underline{\theta}) = L(s,i,\underline{\theta})$ in the sense of Eaton (1967) for all permutations π), and:

(b) For every $\underline{\theta} \in \Omega^k$ and $i,j \in \{1,\ldots,k\}$ with $\theta_i < \theta_j$, $L(\{i\},i,\underline{\theta}) \geq L(\{j\},j,\underline{\theta})$; and $s \subseteq \{1,\ldots,k\}$ with $i,j \in s$ implies $L(s,i,\underline{\theta}) \geq L(s,j,\underline{\theta})$.

The risk of a procedure (S,d) at $\underline{\theta} \in \Omega^k$ is given by

$$(1)\quad r_{\underline{\theta}}(S,d) = E_{\underline{\theta}}(L(S(\underline{U}),\ d_{S(\underline{U})}(\underline{U},\underline{V}),\ \underline{\theta}))$$

$$= L(\emptyset,0,\underline{\theta})P_{\underline{\theta}}\{S(\underline{U}) = \emptyset\}$$

$$+ \sum_{i=1}^{k} L(\{i\},i,\underline{\theta})P_{\underline{\theta}}\{S(\underline{U})=\{i\}\,|\,|S(\underline{u})| = 1\}\ P_{\underline{\theta}}\{|S(\underline{U})|=1\}$$

$$+ \sum_{s,|s|\geq 2} \Big(\sum_{i\in s} L(s,i,\underline{\theta})P_{\underline{\theta}}\{d_s(\underline{U},\underline{V})=i\,|\,S(\underline{U})=s\}\Big)P_{\underline{\theta}}\{S(\underline{U})=s\}.$$

Our first result is with respect to final decisions at Stage 1.

Lemma 1. Let (S,d) be a 2-stage procedure and let $\tilde{S}$ be the same procedure as S with the only modification that for all $\underline{u} \in \mathbb{R}^k$ with $|S(\underline{u})| = 1$,

$\tilde{S}(\underline{u}) = \{i\}$ implies $u_i = \max_{j=1,\ldots,k} u_j$, $i \in \{1,\ldots,k\}$.

Then $r_{\underline{\theta}}(\tilde{S},d) \leq r_{\underline{\theta}}(S,d)$ for all $\underline{\theta} \in \Omega^k$.

Proof: For a fixed (S,d), let $A = \{\underline{u} \in \mathbb{R}^k | \ |S(\underline{u})| = 1\}$. Let $\underline{\theta} \in \Omega^k$ with $P_{\underline{\theta}}\{\underline{U} \in A\} > 0$. The conditional distribution of $\underline{U}$, given $\underline{U} \in A$, has the following density w.r.t. the Lebesgue measure

$$(2) \qquad P_{\underline{\theta}}\{\underline{U} \in A\}^{-1} \prod_{i=1}^{k} f_{\theta_i}(u_i)\, I_A(\underline{u}), \quad \underline{u} \in \mathbb{R}^k.$$

Since by the invariance of S, A is permutation symmetric and moreover, $P_{\underline{\theta}}\{\underline{U} \in A\}$ is a permutation invariant function of $\underline{\theta} \in \Omega^k$, (2) is of the form assumed in Lehmann (1966). Also, $L(\{i\}, i, \underline{\theta})$ satisfies the monotonicity property (5) of Lehmann (1966). Thus by his main result the first sum in (1) is for $(\tilde{S},d)$ smaller or equal to that one for (S,d). Since all other terms in (1) are the same for both procedures the proof is completed.

The corresponding proof with respect to final decisions at Stage 2 uses basically the same idea, but the analysis turns out to be a bit more complicated. For simplicity, let us assume from now on that the mapping $(u,v) \to (u,T(u,v))$ for (u,v) in the interior of $\mathcal{D} = \bigcup_{\theta\in\Omega} (\text{support}(f_\theta) \text{x} \text{support}(g_\theta))$ is one-to-one and continuously differentiable. Thus we have a function $\tilde{T}$ with $(u,v) = (u,\tilde{T}(u,T(u,v)))$, $(u,v) \in \mathcal{D}$ with analogous properties.

Lemma 2. For every $\underline{\theta} \in \Omega^k$ and every permutation symmetric Borel set $A \subseteq \mathbb{R}^k$ with $P_{\underline{\theta}}\{\underline{U} \in A\} > 0$, the conditional distribution of $\underline{W}$, given $\underline{u} \in A$, has a density w.r.t. the Lebesgue measure of the type

$$(3) \qquad c(\underline{\theta}) \prod_{i=1}^{k} \tilde{h}_{\theta_i}(w_i)\, p(\underline{w}), \quad \underline{w} \in \mathbb{R}^k,$$

where c: $\Omega^k \to \mathbb{R}_+ = \{\xi | \xi \geq 0\}$ is permutation invariant, $\tilde{h}_\theta$: $\mathbb{R} \to \mathbb{R}_+$ is measurable, p: $\mathbb{R}^k \to \mathbb{R}_+$ is measurable permutation invariant, and $\{\tilde{h}_\theta\}_{\theta\in\Omega}$ has MLR.

Proof: Let $\underline{\theta} \in \Omega^k$ and $A \subseteq \mathbb{R}^k$ be given as stated above. Then the conditional distribution of $(\underline{U},\underline{V})$, given $\underline{U} \in A$, has the following density w.r.t. Lebesgue measure.

$$(4) \qquad P_{\underline{\theta}}\{\underline{U} \in A\}^{-1} \prod_{i=1}^{k} f_{\theta_i}(u_i) g_{\theta_i}(v_i)\, I_A(\underline{u}), \quad \underline{u},\underline{v} \in \mathbb{R}^k.$$

Since $W_i = T(U_i,V_i)$ is sufficient for θ_i, $i = 1,\dots,k$, by the factorization theorem there exist non-negative measurable functions $\tilde{h}_\theta$ and G with $f_\theta(u)g_\theta(v) = \tilde{h}_\theta(T(u,v))G(u,v)$, $u,v \in \mathbb{R}$, $\theta \in \Omega$. After inserting this into (4) and after a standard change of variables, we see that the conditional distribution of $(\underline{U},\underline{W})$ with $W_i = T(U_i,V_i)$, $i = 1,\dots,k$, given $\underline{U} \in A$, has the density

$$(5) \qquad P_{\underline{\theta}}\{\underline{U} \in A\}^{-1} \prod_{i=1}^{k} \tilde{h}_{\theta_i}(w_i) G(u_i,\tilde{T}(u_i,w_i)) \left|\frac{\partial \tilde{T}(u_i,w_i)}{\partial w_i}\right| I_A(\underline{u}), \quad \underline{u},\underline{w} \in \mathbb{R}^k.$$

Thus by integrating out the variables $\underline{u} \in \mathbb{R}^k$ we see that the conditional distribution of $\underline{W}$, given $\underline{U} \in A$, has a density of the form (3) w.r.t. the Lebesgue measure. Here $c(\underline{\theta}) = P_{\underline{\theta}}\{\underline{U} \in A\}^{-1}$ which is, as we know already, permutation invariant. Moreover,

$$(6) \qquad p(\underline{w}) = \int_A \prod_{i=1}^{k} G(u_i,\tilde{T}(u_i,w_i)) \left|\frac{\partial \tilde{T}(u_i,w_i)}{\partial w_i}\right| d\underline{u}, \quad \underline{w} \in \mathbb{R}^k,$$

which likewise is permutation invariant. Finally, since by assumption the family of densities for $W_i = T(U_i,V_i)$, $i = 1,\dots,k$, $\{h_\theta\}_{\theta\in\Omega}$, has MLR, $\{\tilde{h}_\theta\}_{\theta\in\Omega}$ has also MLR. This completes the proof of Lemma 2.

Corollary 1. For every $\underline{\theta} \in \Omega^k$, permutation symmetric Borel set $A \subseteq \mathbb{R}^k$ with $P_{\underline{\theta}}\{\underline{U} \in A\} > 0$, and $s = \{i_1,\ldots,i_t\} \subseteq \{1,\ldots,k\}$, the conditional distribution of $(W_{i_1},\ldots,W_{i_t})$, given $\underline{U} \in A$, has a density w.r.t. the Lebesgue measure of the type

$$(7) \qquad c(\underline{\theta}) \prod_{j=1}^{t} \tilde{h}_{\theta_{i_j}}(\xi_j) p_{\underline{\theta}'}(\underline{\xi}), \quad \underline{\xi} \in \mathbb{R}^t,$$

where c and $\{\tilde{h}_\theta\}_{\theta\in\Omega}$ are the same as in (3), $\underline{\theta}' = (\theta_{j_1},\ldots,\theta_{j_{k-t}})$ with $\{i_1,\ldots,i_t\} \cup \{j_1,\ldots,j_{k-t}\} = \{1,\ldots,k\}$, and $p_{\underline{\theta}'}(\underline{\xi})$ is permutation invariant in $\underline{\theta}'$ as well as in $\underline{\xi} \in \mathbb{R}^t$.

Proof: This follows from Lemma 2 by integrating out in (3) the variables $w_{j_1},\ldots,w_{j_{k-t}}$. Especially thereby we get for $(w_{i_1},\ldots,w_{i_t}) \in \mathbb{R}^t$

$$(8) \qquad p_{\underline{\theta}'}(w_{i_1},\ldots,w_{i_t}) = \int_{\mathbb{R}^{k-t}} \prod_{r=1}^{k-t} \tilde{h}_{\theta_{j_r}}(w_{j_r}) p(w_{i_1},\ldots,w_{i_t},w_{j_1},\ldots,w_{j_{k-t}}) \, d(w_{j_1},\ldots,w_{j_{k-t}}),$$

which now can be seen to have the invariance properties stated above. Thus the proof is completed.

Now we are ready to prove the main result of this section.

Theorem 1. Let (S,d) be a 2-stage procedure and let $\tilde{S}$ be the modification of S (given in Lemma 1) which uses the optimal final decision at Stage 1. Let $d^* = \{d_s^*\}_{s \subseteq \{1,\ldots,k\}}$ be the set of natural final decisions, i.e. where for every $s \subseteq \{1,\ldots,k\}$ with $|s| \geq 2$, $i \in s$, $\underline{u}, \underline{v} \in \mathbb{R}^k$ and $d_s^*(\underline{u},\underline{v}) = i$ implies $T(u_i,v_i) = \max_{j\in s} T(u_j,v_j)$. Then $r_{\underline{\theta}}(\tilde{S},d^*) \leq r_{\underline{\theta}}(S,d)$ for all $\underline{\theta} \in \Omega^k$.

Proof: Let $\underline{\theta} \in \Omega^k$ be fixed. In view of Lemma 1 we only have to show that $r_{\underline{\theta}}(S,d^*) \leq r_{\underline{\theta}}(S,d)$. Thus by (1) it suffices to prove that for every $s \subseteq \{1,\ldots,k\}$ with $|s| \geq 2$ and $P_{\underline{\theta}}\{S(\underline{U}) = s\} > 0$,

$$\text{(9)} \qquad \sum_{i \in s} L(s,i,\underline{\theta}) P_{\underline{\theta}}\{d_s^*(\underline{u},\underline{v}) = i \,|\, S(\underline{U}) = s\}$$

$$\leq \sum_{i \in s} L(s,i,\underline{\theta}) P_{\underline{\theta}}\{d_s(\underline{U},\underline{V}) = i \,|\, S(\underline{U}) = s\}.$$

Let $s \subseteq \{1,\ldots,k\}$ with $|s| \geq 2$ be fixed. Let $A = \{\underline{u} \in \mathbb{R}^k | S(\underline{u}) = s\}$. By the invariance property of S, A is permutation symmetric. In the conditional situation, given $S(\underline{U}) = s$ or, equivalently, given $\underline{U} \in A$, $\underline{W}$ is sufficient for $\underline{\theta} \in \Omega^k$. This can be seen from (4) and the sentence following (4). Thus, similar as one concludes in the theory of selection procedures, if $s = \{i_1,\ldots,i_t\}$ with $1 \leq i_1 < \ldots < i_t \leq k$, say, then we can assume that $d_s(\underline{u},\underline{v})$ is a function of $(T(u_{i_1},v_{i_1}),\ldots,T(u_{i_t},v_{i_t}))$. By Corollary 1 the conditional distribution of $(W_{i_1},\ldots,W_{i_t})$, given $\underline{U} \in A$, has a density w.r.t. the Lebesgue measure of the form (7) or, respectively,

$$\text{(10)} \qquad c_{\underline{\theta}'}(\theta_{i_1},\ldots,\theta_{i_t}) \prod_{j=1}^{t} \tilde{h}_{\theta_{i_j}}(\xi_j) p_{\underline{\theta}'}(\underline{\xi}), \quad \underline{\xi} \in \mathbb{R}^t,$$

where $c_{\underline{\theta}'}: \Omega^t \to \mathbb{R}_+$ and $p_{\underline{\theta}'}: \mathbb{R}^t \to \mathbb{R}_+$ are permutation invariant functions, $p_{\underline{\theta}'}$ is measurable and $\{\tilde{h}_\theta\}_{\theta \in \Omega}$ has MLR. Therefore this distribution satisfies all conditions assumed by Lehmann (1966). Since moreover, $L(s,i,\underline{\theta})$, $i \in s$, satisfies the condition (5) in his paper, it follows from his version of the "Bahadur-Goodman Theorem" that inequality (9) holds. This completes the proof of the theorem.

Remark 1. Let (S,d) be any 2-stage procedure. Then, after having made a decision S=s, say, the final decision d = i, say, can be viewed as being a partition $(s \setminus \{i\}, \{i\})$ of s into two subsets of sizes $|s|-1$ and 1, respectively. If, more generally, partitions into q subsets of s of fixed sizes $r_1,\dots,r_q$ are to be made, where $q,r_1,\dots,r_q$ depend on $|s|$, then the more general version of the "Bahadur-Goodman-Theorem" by Eaton (1967) can be applied. Thus, if the loss structure in this setting is compatible with the one assumed by Eaton (1967), then the set of natural partitions in terms of the ordered W_i's is optimal.

By Theorem 1 we know now especially, that after having made a decision $S(\underline{U}) = s$, say, it is always better to make a final decision in terms of the largest W_{i_0} among the W_i with $i \in s$, than to make it in terms of the largest V_{i_0} among the V_i with $i \in s$. This fact appears to be interesting enough to be formulated in a slightly more general form in the following Corollary 2.

Corollary 2. For every $\underline{\theta} \in \Omega^k$, $s \subseteq \{1,\dots,k\}$ and every permutation symmetric Borel set $A \subseteq \mathbb{R}^k$ with $P_{\underline{\theta}}\{\underline{U} \subset A\} > 0$ the following holds true. Let $e_i = P_{\underline{\theta}}\{W_i = \max_{j \in s} W_j | \underline{U} \in A\}$ and $f_i = P_{\underline{\theta}}\{V_i = \max_{j \in s} V_j\}$, $i \in s$. Then the e_i's and f_i's are ordered in the same order as the θ_i's with $i \in s$ and, moreover, the vector of e_i's majorizes the vector of f_i's.

Proof: Without loss of generality, let $s = \{1,\dots,t\}$ with $t \geq 2$ and $\underline{\theta} \in \Omega^k$ with $\theta_1 \leq \dots \leq \theta_t$. If $A \subseteq \mathbb{R}^k$ has the properties stated above, take any permutation invariant S with $S(\underline{u}) = s$ if $\underline{u} \in A$ and with $|S(\underline{u})| \leq 1$ otherwise. Let $r \in \{1,\dots,t-1\}$ be fixed and take any loss structure L with $L(s,i,\underline{\theta}) = 1(0)$ if $i \leq (>) r$, $i \in s$. Let $d_s(\underline{u},\underline{v}) = i$ if $v_i = \max_{j \in s} v_j$, $i \in s$, where ties are broken at random. Then by Theorem 1 we get $r_{\underline{\theta}}(S,d^*) \leq r_{\underline{\theta}}(S,d)$

or, more specifically, by inequality (9) we get $f_{r+1} + \ldots + f_t \leq e_{r+1} + \ldots + e_t$, since, obviously, we have $f_1 + \ldots + f_t = e_1 + \ldots + e_t = 1$. Moreover, that $f_1 \leq \ldots \leq f_t$ holds is well known. Finally, $e_1 \leq \ldots \leq e_t$ follows from Corollary 1 and Lemma 4.1 of Eaton. Thus the proof is completed.

Example (Normal Case): Let us look at the special case where $\pi_1, \ldots, \pi_k$ are normal populations $N(\theta_1, \sigma^2), \ldots, N(\theta_k, \sigma^2)$ with unknown means $\theta_1, \ldots, \theta_k \in \mathbb{R}$ and a common known variance $\sigma^2 > 0$. Let U_i and V_i be the arithmetic means of the observations in samples $\{X_{ij}\}_{j=1,\ldots,n}$ and $\{Y_{ij}\}_{j=1,\ldots,m}$, respectively, $i = 1, \ldots, k$. In several parts of this paper we shall return to this special case which henceforth will be denoted as the normal case.

Thus we have $U_i \sim N(\theta_i, p)$ and $V_i \sim N(\theta_i, q)$, $i = 1, \ldots, k$, which are mutually independent, where $p = \sigma^2/n$ and $q = \sigma^2/m$. Let W_i be the overall arithmetic mean for π_i, $i = 1, \ldots, k$. Then $W_i = T(U_i, V_i) = (qU_i + pV_i)/(q+p) \sim N(\theta_i, (q^{-1}+p^{-1})^{-1}\sigma^2)$, and $V_i = \tilde{T}(U_i, W_i) = W_i + p^{-1}q(W_i - U_i)$, $i = 1, \ldots, k$.

Since all our assumptions concerning the underlying distributions are satisfied, all results derived so far are valid in this case. And from Corollary 2, one can derive interesting inequalities.

Remark 2. Without going into details it should be pointed out that analogous results to the ones derived in this section can be obtained in more general sequential settings, provided that the stopping rule is permutation invariant.

3. A natural type 2-stage procedure. In this section we will study 2-stage procedures (S,d) from a non-decision theoretic point of view. Let a correct decision (CD) at $\underset{\sim}{\theta} \in \Omega^k$ be $d = 0$ (i.e. $S = \emptyset$) if $\theta_1,\ldots,\theta_k \leq \theta_0$, and be $d = i$ if $\theta_i = \max_{j=1,\ldots,k} \theta_j > \theta_0$, otherwise. Let us assume that the experimenter wishes to have a procedure (S,d) which at Stage 1 has a small expected number of selected bad populations, denoted by $E_{\underset{\sim}{\theta}}(N_b)$ (a small expected overall sampling amount or a small similar measure of economical performance), and a large probability of a correct selection $P_{\underset{\sim}{\theta}}(CD)$ at points $\underset{\sim}{\theta} \in \Omega^k$ where $\max_{j=1,\ldots,k} \theta_j > \theta_0$, subject to the basic $P_\star$-condition $\inf\{P_{\underset{\sim}{\theta}}(CD) \mid \underset{\sim}{\theta} \in \Omega^k,\ \theta_1,\ldots,\theta_k \leq \theta_0\} \geq P_\star$, where $P_\star$ is a pre-specified constant with $0 < P_\star < 1$.

The following procedure may, sometimes, be applied in practice. The experimenter takes the UMP-test for H: $\theta \leq \theta_0$ versus K: $\theta > \theta_0$ at level $\alpha = 1-P_\star^{1/k}$ and selects all populations π_i with are shown to be significantly good by statistics U_i, $i = 1,\ldots,k$. His final decision may be the natural one based on the V_i's associated with the populations which are selected at Stage 1. From Corollary 2 it follows that this procedure can be improved with respect to $P_{\underset{\sim}{\theta}}(CD)$ without any changes in the expected number of selected good populations $E_{\underset{\sim}{\theta}}(N_g)$, $E_{\underset{\sim}{\theta}}(N_b)$ and $P_{\underset{\sim}{\theta}}\{S(\underset{\sim}{U}) = \emptyset\}$. This procedure ρ will be studied now in more detail. For convenience, let us define it without using the terminology of hypothesis testing.

Definition 3. Procedure ρ. Let ρ be the 2-stage procedure (S,d^*) with $S(\underset{\sim}{u}) = \{i \mid u_i > a,\ i = 1,\ldots,k\}$, where $a \in \mathbb{R}$ is determined by $P_{\theta_0}\{U_1 \leq a\}^k = P_\star$.

That ρ satisfies the basic $P_\star$-condition follows from the fact that U_i is stochastically non-decreasing in $\theta_i \in \Omega$, $i = 1,\ldots,k$, which in turn is a well-known consequence of the MLR property of $\{f_\theta\}_{\theta\in\Omega}$.

In the next two steps we establish formulas for the distribution of final decisions under ρ and derive a basic monotonicity property.

Theorem 2. For every $\underline{\theta} \in \Omega^k$

$$(11) \qquad P_{\underline{\theta}}\{d^*_{S(\underline{U})}(\underline{U},\underline{V}) = i\} = \begin{cases} \prod_{j=1}^{k} P_{\theta_j}\{U_j \le a\} = \prod_{j=1}^{k} F_{\theta_j}(-\infty), & i = 0, \\ \int_{\mathbb{R}} \prod_{\substack{j=1 \\ j \ne i}}^{k} F_{\theta_j}(\lambda) dF_{\theta_i}(\lambda), & i = 1,\dots,k, \end{cases}$$

where for $\theta_r \in \Omega$, $\lambda \in \mathbb{R} \cup \{-\infty\}$, $r = 1,\dots,k$,

$$(12) \qquad F_{\theta_r}(\lambda) = E_{\theta_r}[I_{(-\infty,a]}(U_r) + (1-I_{(-\infty,a]}(U_r))I_{(-\infty,\lambda]}(W_r)].$$

Proof: For $r \in \{1,\dots,k\}$ take the improper random variable defined by $Z_r = -\infty$ (W_r) if $U_r \le$ ($>$)a, which obviously has the distribution function $F_{\theta_r}(\lambda)$, $\lambda \in \mathbb{R} \cup \{-\infty\}$. Now, for $i = 1,\dots,k$ we have

$$(13) \qquad \begin{aligned} &\{Z_i = \max_{j=1,\dots,k} Z_j, \text{ and } Z_i > -\infty\} \\ &= \{W_i = \max\{W_j | U_j > a, \ j=1,\dots,k\}, \text{ and } U_i > a\} \\ &= \{d^*_{S(\underline{U})}(\underline{U},\underline{V}) = i\}. \end{aligned}$$

Therefore, in view of the independence of $Z_1,\dots,Z_k$, (11) follows for $i = 1,\dots,k$. The proof of (11) for $i = 0$ is straightforward.

Theorem 3. $\{F_\theta\}_{\theta \in \Omega}$, as given in (12), is a stochastically non-decreasing family of distribution functions on $\mathbb{R} \cup \{-\infty\}$.

Proof: Let $\lambda \in \mathbb{R} \cup \{-\infty\}$ be fixed and let $H_{a,\lambda}$ be an auxiliary function defined by

(14) $\quad H_{a,\lambda}(u,v) = (1-I_{(-\infty,a]}(u))(1-I_{(-\infty,\lambda]}(T(u,v)), \ (u,v) \in \mathcal{B}.$

By the assumptions made in Section 1 we can assume that $T(u,v)$ is a non-decreasing function in u as well as in v, $(u,v) \in \mathcal{B}$. Thus $H_{a,\lambda}$ has the same monotonicity properties. Now by $W_1 = T(U_1,V_1)$ and (12) we have

(15)
$$F_{\theta_1}(\lambda) = 1-E_{\theta_1}[(1-I_{(-\infty,a]}(U_1))(1-I_{(-\infty,\lambda]}(W_1))]$$
$$= 1-E_{\theta_1}[H_{a,\lambda}(U_1,V_1)], \quad \theta_1 \in \Omega.$$

Since U_1 and V_1 are stochastically non-decreasing in $\theta_1 \in \Omega$ and independent, $E_{\theta_1}[H_{a,\lambda}(U_1,V_1)]$ is non-decreasing in $\theta_1 \in \Omega$. This follows from Lehmann (1955). Thus the proof is completed.

From Theorems 2 and 3 several desirable properties of procedure $\mathcal{P}$ can be derived. Properties 1-4 can be proved with standard techniques (especially integration by parts) from single stage selection theory. The masses in $\{-\infty\}$ have to be taken into consideration, but they cause no serious problems. Thus, we omit the proofs for brevity.

Property 1: For every $i \in \{1,\ldots,k\}$, $P_{\underline{\theta}}\{d^*_{S(\underline{U})}(\underline{U},\underline{V}) = i\}$ is non-decreasing in θ_i and non-increasing in θ_j, $j \neq i$, $\underline{\theta} \in \Omega^k$.

Property 2: For every $\underline{\theta} \in \Omega^k$ with $\theta_1 \leq \ldots \leq \theta_k$, $P_{\underline{\theta}}\{d^*_{S(\underline{U})}(\underline{U},\underline{V}) = i\}$ is non-decreasing in $i \in \{1,\ldots,k\}$.

Property 3: For every non-empty set $M \subseteq \{1,\ldots,k\}$, $P_{\underline{\theta}}\{d^*_{S(\underline{U})}(\underline{U},\underline{V}) \in M\}$ is non-decreasing (non-increasing) in θ_i with $i \in M(i \notin M)$, $\underline{\theta} \in \Omega^k$.

Property 4: $E_{\underline{\theta}}(N_g)$ $(E_{\underline{\theta}}(N_b))$ is non-decreasing (non-increasing) in θ_i with $\theta_i > \theta_0$ $(\theta_i \leq \theta_0)$, $i = 1,\ldots,k$, $\underline{\theta} \in \Omega^k$.

<u>Property 5</u>: For every $\underline{\theta} \in \Omega^k$ with $\theta_1,\ldots,\theta_k < \theta_0$, $P_{\underline{\theta}}\{S(\underline{U}) = \emptyset\}$ tends to 1 for large n. For every $\underline{\theta} \in \Omega^k$ with $\theta_1,\ldots,\theta_{k-t} < \theta_0 < \theta_{k-t+1},\ldots,\theta_{k-1} < \theta_k$, $t \in \{1,\ldots,k\}$, $P_{\underline{\theta}}\{S(\underline{U}) = \{k-t+1,\ldots,k\}, d^*_{S(\underline{U})}(\underline{U},\underline{V}) = k\}$ tends to 1 for large n and m.

<u>Proof</u>: The first assertion follows from the well known consistency properties of the UMP-test mentioned at the beginning of this section. For $\underline{\theta} \in \Omega^k$ with $\theta_1,\ldots,\theta_{k-t} < \theta_0 < \theta_{k-t+1},\ldots,\theta_{k-1} < \theta_k$, $t \in \{1,\ldots,k\}$, by the same reasons, $P_{\underline{\theta}}\{S(\underline{U}) = \{k-t+1,\ldots,k\}\}$ tends to 1 for large n. Since, moreover, $P_{\underline{\theta}}\{W_k = \max_{j=1,\ldots,k} W_j\}$ tends to 1 for large n+m (see Miescke (1979a)), the proof is completed.

Next we like to show along the lines of Tamhane and Bechhofer, in an informal way of proof, that procedure $\mathcal{P}$ is preferable to the corresponding 1-stage procedure $\mathcal{P}_0$, say, from an economical point of view. Let $\tilde{U}_1,\ldots,\tilde{U}_k$ be of the same type as $U_1,\ldots,U_k$, but based on samples of size n_0 from $\pi_1,\ldots,\pi_k$. Then $\mathcal{P}_0$ decides as follows:

$$(16) \qquad S_0(\tilde{\underline{U}}) = \begin{cases} \emptyset & \text{if } \tilde{U}_1,\ldots,\tilde{U}_k \le a_0 \\ \{i\} & \text{if } \tilde{U}_i = \max_{j=1,\ldots,k} \tilde{U}_j \text{ and } \tilde{U}_i > a_0,\ i=1,\ldots,k, \end{cases}$$

where a_0 is determined by $P_{\theta_0}\{\tilde{U}_1 \le a_0\}^k = P_*$. The version of $\mathcal{P}_0$ in the <u>normal case</u> was studied by Bechhofer and Turnbull (1978). Now, if an optimal allocation of observations is derived subject to a criterion which can be met by the use of monotonicity properties of $P_{\underline{\theta}}\{$final decision is "i"$\}$ in $\theta_1,\ldots,\theta_k$, $i = 1,\ldots,k$, then the allocation problem has to be solved for both, $\mathcal{P}$ and $\mathcal{P}_0$, at the same points $\underline{\theta} \in \Omega^k$. Since then $\mathcal{P}_0$ can be viewed to be a special version of $\mathcal{P}$ with $n = n_0$ and $m = 0$, we conclude as follows:

Property 6: In every allocation problem subject to a criterium which can be met by the use of monotonicity properties of $P_{\underline{\theta}}\{$final decision is "i"$\}$ in $\theta_1,\ldots,\theta_k$, $\underline{\theta} \in \Omega^k$, $i = 1,\ldots,k$, $\mathcal{P}$ is at least as economical as $\mathcal{P}_0$.

Finally, let us consider the class $\mathcal{C}$ of procedures which are of the same type as $\mathcal{P}$ but use another level α test at Stage 1. Then by the properties of the UMP level α test we get

Property 7: For every fixed n, m and α (or P_*, respectively), $\mathcal{P}$ maximizes (minimizes) $E_{\underline{\theta}}(N_g)$ $(E_{\underline{\theta}}(N_b))$ within the class $\mathcal{C}$, uniformly in $\underline{\theta} \in \Omega^k$.

To summarize the results so far derived, and especially in view of Properties 4, 5 and 7, $\mathcal{P}$ appears to be a reasonable procedure if the experimenter wishes to screen out the bad populations at Stage 1, to keep the good ones (if there are any) at the same time, and finally to select the best population (if it is good).

On the other hand, let us look at the case where the experimenter is looking for the best population (if it is good) but wishes to keep the expected overall sampling amount small. Then at points $\underline{\theta} \in \Omega^k$ where more than one population is good, $\mathcal{P}$ might possibly not very effectively screen out. Here an additional screening mechanism seems to be appropriate, i.e. a subset selection procedure for the first stage, which has to be combined with a procedure of the type S considered so far.

In the normal case, analogous to what Tamhane and Bechhofer (1979) proposed in the non-control setting, a natural choice for the additional screening mechanism could be Gupta's (1956) maximum means procedure. This leads to a 2-stage procedure $\mathcal{P}_\Delta = (S_\Delta, d^*)$ with

(17) $\quad S_\Delta(\underline{u}) = \{i \mid u_i > a_\Delta \text{ and } u_i \geq \max_{j=1,\ldots,k} u_j - p^{1/2}\Delta,\ i = 1,\ldots,k\}$,

where $\Delta \geq 0$ is fixed and a_Δ has to be determined such that S_Δ meets the basic P_*-condition. Note that for $\Delta = 0$ (∞), $\mathcal{P}_\Delta$ is of the type $\mathcal{P}_0$ $(\mathcal{P})$.

Since we again have enlarged the class of 2-stage procedures, we are led to a more economical type of procedure in the sense of Property 6. Moreover, for $\Delta > 0$ and $\underline{\theta} \in \Omega^k$ with $\theta_0 < \max_{j=1,\ldots,k} \theta_j$, the probability of making a correct final decision at Stage 1 already, tends to 1 for large n. But, on the other hand, $\mathcal{P}_\Delta$ for $0 < \Delta < \infty$ is much more difficult to implement in practice. The problems arising here are of the same type as discussed in Tamhane and Bechhofer (1979), Gupta and Miescke (1980) and Miescke and Sehr (1980).

4. Bayesian 2-stage procedures. From now on let us assume that the parameters $\underline{\theta} = (\theta_1,\ldots,\theta_k)$ vary randomly according to a permutation invariant prior distribution τ on the Borel sets of Ω^k. We will study the form of Bayesian 2-stage procedures under a loss structure L given by

$$(18)\quad L(s,i,\underline{\theta}) = \begin{cases} 0 & \text{if } s = \emptyset \\ \ell(\theta_0-\theta_i) & \text{if } s = \{i\} \\ c|s|+\ell(\theta_0-\theta_i) & \text{if } |s| \geq 2 \end{cases}$$

$i = 1,\ldots,k$, $\underline{\theta} \in \Omega^k$, where $c \geq 0$ is a constant and $\ell\colon \mathbb{R} \to \mathbb{R}$ is non-increasing, integrable, with $\ell(0) = 0$. The overall Bayes risk is given by

$$(19)\quad \int_{\Omega^k} [\sum_{i=1}^{k} \ell(\theta_0-\theta_i)P_{\underline{\theta}}\{S(\underline{U}) = \{i\}\} + \sum_{s,|s|\geq 2} \Big(c|s| +$$

$$+ \sum_{i\in s} \ell(\theta_0-\theta_i)P_{\underline{\theta}}\{d_s(\underline{U},\underline{V})=i\,|\,S(\underline{U})=s\}\Big)P_{\underline{\theta}}\{S(\underline{U})=s\}]d\tau(\underline{\theta}).$$

By Theorem 1 we can restrict our consideration to Bayes procedures (S^B,d^B) with $d^B = d^*$ and the property that $\underline{u} \in \mathbb{R}^k$ and $S^B(\underline{u}) = \{i\}$ implies $u_i = \max_{j=1,\ldots,k} u_j$, $i = 1,\ldots,k$. Therefore at every point $\underline{u} \in \mathbb{R}^k$ an optimal subset selection procedure S^B decides in favor of a subset $s \subseteq \{1,\ldots,k\}$ which is associated with the smallest of the values given in the following scheme.

s	Posterior risk at $\underline{u} \in \mathcal{S}^k$, $\mathcal{S} = \bigcup_{\theta\in\Omega}$ support (f_θ)
$\emptyset$	0
$\{i\}$	$E\{\ell(\theta_0-\Theta_i)\,|\,\underline{U}=\underline{u}\}$, $u_i = \max_{j=1,\ldots,k} u_j$
$\{i_1,\ldots,i_t\}$	$tc + E\{\min_{j=1,\ldots,t} E\{\ell(\theta_0-\Theta_{i_j})\,|\,\underline{U}=\underline{u},\, \underline{V}\}\ \|\ \underline{U}=\underline{u}\}$, $1 \leq i_1<\ldots< i_t \leq k$, $t \geq 2$.

Note that in the last expression the inner conditional expectation is viewed as being a function of $\underline{V}$, and that the outer one is the expectation with respect to the conditional distribution of $\underline{V}$, given $\underline{U}=\underline{u}$.

Definition 4. A 2-stage procedure (S,d) is called monotone if $(S,d) = (\tilde{S},d^*)$ in the sense of Theorem 1 and, moreover, if for every $\underline{u} \in \mathbb{R}^k$, $i,j \in \{1,\ldots,k\}$ with $u_i < u_j$, $i \in S(\underline{u})$ implies $j \in S(\underline{u})$.

Next we wish to find sufficient conditions under which there exist Bayes 2-stage procedures which are monotone. For this purpose let $\underline{u} \in \mathscr{S}^k$ with $u_1 \leq \ldots \leq u_k$ and $t \in \{2,\ldots,k-1\}$ be fixed. In Goel and Rubin (1977) an optimal s with $|s| = t$ could be derived directly from Eaton's result. In our situation this is not possible since now the conditional expected loss, given $\underline{U} = \underline{u}$, does not simply depend on $S(\underline{u})$, but also on $\underline{u}$. Let us now try to find sufficient conditions under which the posterior risk at $\underline{u}$ is minimal for the set $\{k-t+1,\ldots,k\}$ among all sets $s \subseteq \{1,\ldots,k\}$ with $|s| = t$. An optimal s with $|s| = t$ minimizes

$$(20) \qquad E\{\min_{j\in s} E\{\ell(\theta_0-\theta_j)|\underline{U} = \underline{u}, \underline{V}\} \mid \underline{U} = \underline{u}\}$$

$$= \int_{\mathbb{R}^k} \min_{j\in s} \int_{\Omega^k} \ell(\theta_0-\theta_j) \prod_{i=1}^{k} f_{\theta_i}(u_i) g_{\theta_i}(v_i) d\tau(\underline{\theta}) d\underline{v}\ \beta(\underline{u}),$$

where $\beta(\underline{u}) = (\int_{\Omega^k} \prod_{i=1}^{k} f_{\theta_i}(u_i) d\tau(\underline{\theta}))^{-1}$ is of no relevance for our problem and thus can be ignored in the sequel. From the remark following (4) we see that the integral on the r.h.s. of (20) can be rewritten as

$$(21) \qquad \int_{\mathbb{R}^k} \min_{j\in s} \int_{\Omega^k} \ell(\theta_0-\theta_j) \prod_{i=1}^{k} \tilde{h}_{\theta_i}(T(u_i,v_i)) d\tau(\underline{\theta}) \prod_{r=1}^{k} G(u_r,v_r) d\underline{v}.$$

A change of variables $w_i = T(u_i,v_i)$ (or $v_i = \tilde{T}(u_i,w_i)$, respectively) modifies (21) to

$$(22)\qquad \int_{\mathbb{R}^k} \min_{j\in s} \int_{\Omega^k} \ell(\theta_0-\theta_j) \prod_{i=1}^{k} \tilde{h}_{\theta_i}(w_i)d\tau(\underline{\theta}) \prod_{r=1}^{k} G(u_r,\tilde{T}(u_r,w_r)) \left|\frac{\partial\tilde{T}(u_r,w_r)}{\partial w_r}\right| d\underline{w}.$$

Now we are in position to apply Eaton's main result iteratively, first to the inner integral (i.e. the 2nd stage scenario), and then to the outer one (i.e. the 1st stage scenario). Let L_s: $\mathbb{R}^k \to \mathbb{R}$ be defined by

$$(23)\qquad L_s(\underline{w}) = \min_{j\in s} \int_{\Omega^k} \ell(\theta_0-\theta_j) \prod_{i=1}^{k} \tilde{h}_{\theta_i}(w_i)d\tau(\underline{\theta}), \quad \underline{w} \in \mathbb{R}^k.$$

Lemma 3. For every $\underline{w} \in \mathbb{R}^k$, $s \subseteq \{1,\ldots,k\}$ with $|s| = t$, $i \in s$, $j \in \{1,\ldots,k\}\setminus s$, $\tilde{s} = (s\setminus\{i\}) \cup \{j\}$ and $w_i \le w_j$ implies $L_{\tilde{s}}(\underline{w}) \le L_s(\underline{w})$.

Proof: Let $r \in \{1,\ldots,k\}$ and $\underline{w} \in \mathbb{R}^k$ be fixed. Then except for a normalizing factor depending on $\underline{w}$, $R_r = \int_{\Omega^k} \ell(\theta_0-\theta_r) \prod_{i=1}^{k} \tilde{h}_{\theta_i}(w_i)d\tau(\underline{\theta})$ can be viewed as being the posterior risk ($\underline{w}$ are the given "observations" and $\underline{\theta}$ are the "parameters") for decision $\{r\}$ in a fixed size 1 subset selection problem of the type treated in Eaton (1967). The loss function hereby is $\tilde{L}_{\{i\}}(\underline{\theta}) = \ell(\theta_0-\theta_i)$, $\underline{\theta} \in \Omega^k$, $i = 1,\ldots,k$, which clearly satisfies the monotonicity and invariance properties (3.4) and (3.5) of Eaton (1967). Thus by his Lemma 4.1 we know that $R_1,\ldots,R_k$ are ordered in the reverse order to $w_1,\ldots,w_k$. This completes the proof.

In view of Lemma 3 we know now that an optimal s with $|s| = t$ minimizes

$$(24)\qquad \int_{\mathbb{R}^k} L_s(\underline{w}) \prod_{i=1}^{k} G(u_i,\tilde{T}(u_i,w_i)) \left|\frac{\partial\tilde{T}(u_i,w_i)}{\partial w_i}\right| d\underline{w},$$

which can be viewed to be (except for a normalizing factor depending on $\underline{u}$) the posterior risk ($\underline{u}$ are the "observations" and $\underline{w}$ the "parameters") for decision s in a fixed size t subset selection problem of the type treated in Eaton (1967). The loss function $L_s(\underline{w})$ hereby satisfies (by Lemma 3) the monotonicity property (3.4) and, obviously, also the invariance property (3.5) of Eaton (1967). Thus by his Lemma 4.1 we see that the following Assumption (A) is sufficient for the existence of a monotone optimal s with $|s| = t$.

Assumption (A). The distributions are as stated in Section 1, and the function $G(u,\tilde{T}(u,w)) \left.\frac{\partial \tilde{T}(\xi,\eta)}{\partial \eta}\right|_{(\xi,\eta)=(u,w)}, (u,w) \in \{(u,T(u,v)) | (u,v) \in \mathcal{D}\}$, has MLR.

Theorem 4. Under Assumption (A), for every loss structure L of type (18) and every permutation invariant prior distribution τ, there exists a 2-stage Bayes procedure (S^B,d^B) which is monotone.

It is now of interest to find simple sufficient conditions for Assumption (A) to hold true. For exponential families we get the following.

Assumption (B).

The underlying distributions for all observations from $\pi_1,\ldots,\pi_k$ belong to an exponential family with densities $a(\theta)b(x)\exp\{\theta x\}$, $x \in \mathbb{R}$, $\theta \in \Omega$, w.r.t. the Lebesgue measure on $\mathbb{R}$, where the function $b(x)$, $x \in \mathbb{R}$, is log-concave (i.e. the densities are strongly unimodal.)

Theorem 5. Assumption (B) implies Assumption (A).

Proof: Let $U_i = \sum_{j=1}^{n} X_{ij}$, $V_i = \sum_{j=1}^{m} Y_{ij}$ and $W_i = U_i+V_i$, $i = 1,\ldots,k$.

Thus we have $T(u,v) = u+v$ and $\tilde{T}(u,w) = w-u$, $u,v,w \in \mathbb{R}$. The density of U_i is $a(\theta_i)^n b^{*n}(u)\exp\{\theta_i u\}, u \in \mathbb{R}$, and the density of V_i is $a(\theta_i)^m b^{*m}(v)$ $\exp\{\theta_i v\}$, $v \in \mathbb{R}$, $i = 1,\ldots,k$, where $b^{*n}(b^{*m})$ denotes the n-fold (m-fold) convolution of b with respect to the Lebesgue measure on $\mathbb{R}$. It follows that

$$(25) \qquad G(u,\tilde{T}(u,w)) \left.\frac{\partial\tilde{T}(\xi,\eta)}{\partial\eta}\right|_{(\xi,\eta)=(u,w)} = b^{*n}(u)b^{*m}(w-u),\ u,w \in \mathbb{R}.$$

Let the function $b(x)$, $x \in \mathbb{R}$, be log-concave. Then by Ibragimov (1956), the function $b^{*m}(x)$, $x \in \mathbb{R}$, has the same property. But this is equivalent for $b^{*m}(w-u)$ to have MLR in $w \in \mathbb{R}$ w.r.t. $u \in \mathbb{R}$ (cf. Lehmann (1959), p. 330), and therefore it is also equivalent for Assumption (A) to hold true.

Remark 3. It is not difficult to see that in the general case the following conditions are sufficient for Assumption (A) to hold true: $T(u,v) = \varepsilon_1 u + \varepsilon_2 v$, $u,v \in \mathbb{R}$, $\varepsilon_1,\varepsilon_2 > 0$, and $\{g_\theta\}_{\theta\in\Omega}$ is a family of log-concave (i.e. strongly unimodal) densities. This follows directly from the factorization identity $f_\theta(u)g_\theta(v) = \tilde{h}_\theta(T(u,v))G(u,v)$, $u,v \in \mathbb{R}$, $\theta \in \Omega$.

For the remainder of this section let us consider the normal case in more detail. Here, Assumption (B) is satisfied with $b(x) = \exp\{-x^2/2\sigma^2\}$, $x \in \mathbb{R}$, and thus Theorem 4 is valid in this case. Let us assume that apriori $\Theta_1,\ldots,\Theta_k$ are independently identically distributed random normals with mean θ_0 and variance $r > 0$. Then at $\underline{u} \in \mathbb{R}^k$ with $u_1 \le \ldots \le u_k$ the optimal procedure selects at Stage 1 in favor of the smallest value in the following scheme.

s	Posterior risk at $\underset{\sim}{u} \in \mathbb{R}^k$ with $u_1 \leq \ldots \leq u_k$
$\emptyset$	0
$\{k\}$	$E^0[\ell(\alpha_2(\theta_0-u_k) + \alpha_1 Q_0)]$
$\{k-t+1,\ldots,k\}$	$tc + E[\min_{j \geq k-t+1} E^0(\ell(\alpha_2(\theta_0-u_j) + \alpha_3 Q_j + \alpha_4 Q_0))]$, $t \geq 2$,

where $Q_0, Q_1, \ldots, Q_k$ are auxiliary random variables which are independent standard normals, E^0 denotes expectation w.r.t. Q_0, and

$\alpha_1 = (rp(p+r)^{-1})^{1/2}$, $\alpha_2 = r(p+r)^{-1}$, $\alpha_3 = pr[(p+r)(pq+pr+qr)]^{-1/2}$,

$\alpha_4 = [rpq/(pq+pr+qr)]^{1/2}$.

Similar to what was done by Goel and Rubin (1977), let us show next that the Bayes solution at $\underset{\sim}{u} \in \mathbb{R}^k$ with $u_1 \leq \ldots \leq u_k$ can be determined in the following short way. Let r_t denote the posterior risk for decision $s = \{k-t+1,\ldots,k\}$, $i = 0,1,\ldots,k$. At first one compares r_0, r_1, r_2. If $r_0(r_1)$ is the minimum then the final decision is $s = \emptyset$ ($\{k\}$). If $r_2 < r_0$, r_1 then r_2, $r_3, \ldots$ are successively computed until $r_{i_0} \leq r_{i_0+1}$ occurs for the first time; then $s=\{k-i_0+1,\ldots,k\}$ is the final decision. This method is justified by the following result.

<u>Lemma 4. Let</u> $\underset{\sim}{u} \in \mathbb{R}^k$ <u>with</u> $u_1 \leq \ldots \leq u_k$ <u>be fixed and let</u> r_0, $r_1, \ldots, r_k$ <u>be defined as stated above. Then</u> $r_2 - r_3 \geq r_3 - r_4 \geq \ldots \geq r_{k-1} - r_k$.

<u>Proof</u>: Let $Z_1,\ldots,Z_k$ be random variables defined by $Z_j = -E^0(\ell(\alpha_2(\theta_0-u_j) + \alpha_3 Q_j + \alpha_4 Q_0))$, $i = 1,\ldots,k$. Then by $u_1 \leq \ldots \leq u_k$ and the fact that ℓ is non-increasing we have $Z_1 \underset{st}{\leq} Z_2 \underset{st}{\leq} \cdots \underset{st}{\leq} Z_k$ (where "$\underset{st}{\leq}$" denotes stochastic ordering). Moreover, $r_t = tc - E(\max_{j \geq k-t+1}(Z_j))$, $t = 2,\ldots,k$. Thus, for $t \geq 2$, by Chernoff and Yahav (1977) we get

$$r_t - r_{t+1} = E(\max_{j\geq k-t} Z_j) - E(\max_{j\geq k-t+1} Z_j) - c$$

$$= \int_{\mathbb{R}} \prod_{j\geq k-t+1} P\{Z_j \leq \lambda\} P\{Z_{k-t} > \lambda\} d\lambda - c,$$

which clearly is non-increasing in t, $t = 2,\ldots,k-1$.

Let us finally take a brief look at the special case of a linear loss function $\ell(\xi) = a\xi$, $\xi \in \mathbb{R}$, $a > 0$, where we can choose $a = 1$ (since other values of a can be compensated in the cost c). Then the decision at Stage 1 is based on the following scheme.

s	Posterior risk at $\underline{u} \in \mathbb{R}^k$ with $u_1 \leq \ldots \leq u_k$
$\emptyset$	0
$\{k\}$	$\alpha_2(\theta_0-u_k)$
$\{k-t+1,\ldots,k\}$	$\alpha_2(\theta_0-u_k)+tc - \alpha_2 E(\max_{j\geq k-t+1} (u_j-u_k+\alpha_3\alpha_2^{-1}Q_j))$, $t \geq 2$.

Lower and upper bounds for the expectations in this scheme can be found in Miescke (1979b) to approximate the Bayes procedure. Note that if for a $t \in \{2,\ldots,k\}$ $tc \geq \alpha_3 E(\max_{j=1,\ldots,t} Q_j)$, then at most t-1 populations are selected at the first stage. Thus in the case of $2c \geq \alpha_3\pi^{-1/2}$ the Bayes-procedure is of the type ρ_0 (cf. (16)). And for the case of $k = 2$ populations the Bayes-procedure is of the type ρ_Δ (cf. (17)), except for an area in the neighborhood of (θ_0,θ_0) where the Bayes-procedure selects both populations.

Acknowledgment. The authors are grateful to Professors H. Rubin and A. Rukhin for several helpful comments.

REFERENCES

[1] R. E. Bechhofer and B. W. Turnbull (1978). Two (k+1)-decision selection procedures for comparing k normal means with a specified standard. J. Amer. Statist. Assoc., 73, 385-392.

[2] H. Chernoff and J. A. Yahav (1977). A subset selection problem employing a new criterion. Statistical Decision Theory and Related Topics II. (S. S. Gupta and D. S. Moore eds.). Academic Press, 93-119.

[3] M. L. Eaton (1967). Some optimum properties of ranking procedures. Ann. Math. Statist., 38, 124-137.

[4] P. K. Goel and H. Rubin (1977). On selecting a subset containing the best population - a Bayesian approach. Ann. Statist., 5, 969-983.

[5] S. S. Gupta (1956). On a decision rule for a problem in ranking means. Mimeo. Ser. No. 150, Inst. of Statist., Univ. of North Carolina, Chapel Hill.

[6] S. S. Gupta and K. J. Miescke (1980). On the least favorable configurations in certain two-stage selection procedures. Mimeo. Ser. No. 79-6, Dept. of Statist., Purdue Univ., W. Lafayette. To appear in the volume in honor of C. R. Rao.

[7] I. A. Ibragimov (1956). On the composition of unimodal distributions. Theory Prob. Its Appl., 1, 255-260.

[8] E. L. Lehmann (1955). Ordered families of distributions. Ann. Math. Statist., 26, 399-419.

[9] E. L. Lehmann (1959). Testing Statistical Hypotheses. J. Wiley.

[10] E. L. Lehmann (1966). On a theorem of Bahadur and Goodman. Ann. Math. Statist., 37, 1-6.

[11] K. J. Miescke (1979a). Identification and selection procedures based on tests. Ann. Statist., 7, 207-219.

[12] K. J. Miescke (1979b). Bayesian subset selection for additive and linear loss functions. Commun. Statist., A8, 1205-1226.

[13] K. J. Miescke and J. Sehr (1980). On a conjecture concerning the least favorable configuration in certain two-stages selection procedures. Commun. Statist., A9, 1609-1617.

[14] A. C. Tamhane and R. E. Bechhofer (1979). A two-stage minimax procedure with screening for selecting the largest normal mean (II): an improved PCS lower bound and associated tables. Commun. Statist., A8, 337-358.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Mimeograph Series #81-6	2. GOVT ACCESSION NO. AD-A099 613	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) On the Problem of Finding a Best Population with Respect to a Control in Two Stages		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER Mimeo. Series #81-6
7. AUTHOR(s) Shanti S. Gupta and Klaus-J. Miescke		8. CONTRACT OR GRANT NUMBER(s) ONR N00014-75-C-0455
9. PERFORMING ORGANIZATION NAME AND ADDRESS Purdue University Department of Statistics West Lafayette, IN 47907		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Washington, DC		12. REPORT DATE March 1981
		13. NUMBER OF PAGES 25
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

17. DISTRIBUTION ST. [illegible] ENT (of t[illegible] abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY [illegible] TES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Multiple comparison with a control, two-stage selection procedures, screening procedures, sequential procedures, Bayesian procedures, optimal selection.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Let $\pi_1,\ldots,\pi_k$ be given populations associated with unknown real parameters $\theta_1,\ldots,\theta_k$. π_i is assumed to be "good" if $\theta_i > \theta_0$, where $\theta_0 \in \mathbb{R}$ is a given control value, $i = 1,\ldots,k$. The goal is to find the "best" population (i.e. that one with the largest parameter), if it is "good", in 2-stages with screening out "bad" populations in the first stage. Consideration is restricted to permutation invariant procedures. It is shown that under MLR and a general invariant loss structure the natural final decisions are optimum. More generally an extention

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

of the "Bahadur-Goodman Theorem" to sequential settings (with and without relation to a control) is derived. If the loss structure consists of the cost for sampling plus the loss for final decision, it is shown that for every symmetric prior there exists a Bayes procedure which selects at the first stage populations according to the largest observations. Natural procedures, which screen out with the UMP test for H: $\theta \leq \theta_0$ versus K: $\theta > \theta_0$ at fixed level α, are considered. As an example, all results are studied in the special case of normal populations with unknown means and a common known variance.

EN

DATE

ILME